

THE QUALITY OF ENGLISH LANGUAGE TESTING IMPLEMENTED IN KBRI SCHOOL, SEKOLAH INDONESIA KUALA LUMPUR, MALAYSIA

Nur Kholilah

Email: nur_kholilah94@gmail.com

Universitas Islam Negeri Sunan Ampel Surabaya

Abstract. This study is aimed to know how the quality of second grade senior high school Sekolah Indonesia Kuala Lumpur Malaysia English language testing like in term of content validity, index of difficulty, index of discrimination, and the effectiveness of distracters. The design used in this study is a qualitative research. Qualitative in this research is descriptive research. Then, use quantitative descriptive research to calculate and compute the data to prove the qualitative data and conclude the result of this research. The object of this research is second grade of senior high school of Sekolah Indonesia Kuala Lumpur Malaysia and only focus in multiple choice test items. The sample of this research is second grade of Science class and second grade of Social class which conclude all students in the class. The result of this study reported that the English language testing has good content validity. It also reported that the index of difficulty of the English language testing are acceptable. Besides, the index of discrimination of this test is satisfactory. Moreover, it has good distracters.

Key Words: Content Validity, Item analysis, Index of Difficulty, Index of Discrimination, the Effectiveness of Distracters

INTRODUCTION

English learning aimed in junior high school is oriented to reach functional level. It means that the students should be able to communicate oral and written in their daily life activity. While, English learning in senior high school is expected to reach informational level, because they have been prepared to continue their study in university (Depdiknas ; 2004). So, English subject in senior high school is important subject for students have been prepared for their study in university afterward. Senior high school students

expected to master English subject well before they going to university.

Test is used to provide information concerning not only with the individual students performance, but also with the effectiveness of teaching learning activities. And test is one type of measurement is used to measure student's behavior goal of instructions. For teachers, a test is used to measure the effectiveness of teaching learning activities (Mursyidah: 2009).

(Norris: 2000) Language teachers are often faced with the responsibility of selecting or developing language tests for their classrooms and programs. However, deciding which testing alternatives are the most appropriate for a particular language education context can be daunting, especially given the increasing variety of instruments, procedures, and practices available for language testing. Such alternatives include not only test types with long traditions of use—such as multiple choice, matching, true-false, and fill-in-the-blank tests; cloze and dictation procedures; essay exams; and oral interviews—but also tests differing in scope and structure from these well-known options. For example, technological developments have led to a number of new language testing formats, including computer-based and computer-adaptive tests (Brown 1997; Dunkel 1999; Yao and Ning 1998), audiotape-based oral proficiency interviews (Norris 1997; Stansfield and Kenyon 1992), and web-based testing (Roever 1998).

In teaching learning activities, testing has an important role. The results of teaching without testing will be useless, because testing helps to show the achievement of the objectives of education. From the result of the test it can be seen whether the teaching learning process is successful or not. Both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with other (Heaton; 1988). It was cleared that relation between testing and teaching can't be ignored. Teachers, students, and school want to know their effort to achieve the educational objectives are successful or not. They will be satisfied if their efforts are successful. But if their efforts unsuccessful so they will changes their ways (Utami:2013).

Chittenden said that the purpose of testing are “keeping track, checking-up, finding-out, and summing up”. Keeping track is collecting the data about student progress in learning process in the school. Checking-up is checking the students’ skill in learning process and to know weakness of the student in learning process. Finding-out is searching, finding, and detecting the weakness and mistakes from the students in learning process. Summing up is concluding the students’ learning progress which appropriate with standard competency in that school (Arifin: 2012).

Thus, as one tool of evaluation test is needed to be employed in teaching activities. Moreover it has lot of benefits in order to support the success of teaching learning process, such as: (1) To measure language proficiency. (2) To diagnose student’s strengths and weakness, to identify what they know and what they do not know. (3) To discover how successful student have been in achieving the objectives a course of study. (4) To assist placement of student by identifying the stage or part of a teaching program most appropriate to their ability (Hughes: 2003).

Regarding to the case above, it is very important to have tests or some kind another, are valid, well designed and formulated. Hughes mentioned in his book that test is said to be valid if it is measure accurately what it should be measured. Nurkanca and Sumartana also pointed out that a qualified test should be reliable, valid and having degrees of difficulty-index and discriminating power (Nurkanca and Sumartana: 1986).

Language testers are sometimes asked to say what is ‘the best tests’ or ‘the best testing technique’. Such question reveals a misunderstanding of what is involved in the practice of language testing. A test that proves ideal for one purpose maybe quite useless for another; a technique that may work very well in one situation can entirely inappropriate in another. Equally, two teaching institutions may require different test, depending on objectives of their courses, the purpose of the tests, and the resources available (Hughes: 2003). From that point, the teacher must recognize which test that is appropriate to measure the student skills. The teacher must create the test that is suitable with the student ability too.

In this research, the researcher will focus on language testing technique that other the teachers do in schools. The researcher wants to identify how the teacher built the test to the students: what technique that is used by the teachers and how the test can measure the skills of the students and whether the test is suitable with the students or not.

Indonesia has embassy in every country that have relation with Indonesia that called by Embassy of The Republic Indonesia or KBRI (Kedutaan Besar Republik Indonesia). KBRI build school in those countries like Singapore, Malaysia, Thailand etc. which still under KBRI control. One of the examples is Sekolah Indonesia Kuala Lumpur.

Sekolah Indonesia Kuala Lumpur is KBRI's School that is located in Lorong (street) Tun Ismail no.1 50480 Kuala Lumpur Malaysia. Sekolah Indonesia Kuala Lumpur is under control supervision of Indonesians' embassy, it means that the curriculum and the rules of the school are based on Indonesian curriculum.

Like Indonesian school, Sekolah Indonesia Kuala Lumpurs' curriculum is based on Education National Standard or BSNP (Badan Standar Nasional Pendidikan) in Indonesia. From the method term, the teacher use CTL (Contextual Teaching Learning) in English subject. This method aims to help the student to know and use the language in a real situation of the target language. For the textbook of English subject in Senior High School, the teacher uses Indonesian books from Dinas Pendidikan Indonesia or Indonesian Education Agency and Singaporean books. From the method CTL that they use, means they use KTSP or School Based Curriculum. From the competency standards of this school is Standar Kompetensi 2006 as same as in Indonesia.

The reason of the researcher to do this research is, the researcher wants to identify the validity of the test in that school. Also are the culture of Malaysia influenced the teacher on the way they teach the subject and build the test of the students? From those points above, the researcher wants to know how the teachers do a testing for the student in that school and what technique that the teacher use to do a test.

Based on the above background, this research attempts to answer the questions on the quality of the English Language testing used in KBRI school, Sekolah Indonesia, in Malaysia, as well as to describe it.

TESTING AND TEACHING

Test is set techniques, procedures, and items that constitute an instrument of some sort that require performance or activity on the part of the test taker (and sometimes on the part of the tester as well) (Douglas: 2001). Test is procedures designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual (Bachman: 1990).

In line of that, test as quoted from Webster's Collegiate by Daryanto, is any series of questions or exercise or other means of measuring the skill, knowledge, intelligence, capacities of aptitudes or an individual or group (Daryanto: 1999).

In the other word, Kubizyn and Borich stated in their book(2003), that test is just as tools that can contribute importantly to the process of evaluating pupils, the curriculum, and the teaching method.

The effect of testing on teaching learning is known as backwash, and can be harmful or beneficial. If a test is regarded as important, if the stakes are high, preparation for it can come to dominate all teaching and learning activities. And if the test content and testing techniques are at variance with the objective of the course, there is likely to be harmful backwash. An instance of this would be where students are following an English course that is meant to train them in language skills (including writing) necessary for university study in an English speaking country, but where the language test that they have to take in order to be admitted to a university does not test those skills directly. If the skill of the writing, for the example, is tested by multiple choice items, then there is great pressure to practice such items rather that practice the skills of writing itself. This is clearly undesirable (Hughes: 2003).

Standards in testing

One area of increasing concern in language testing has been that of standards. The word 'standards' has various meanings in the literature, as the Task Force on Language Testing Standards set up by ILTA discovered. One common meaning used by respondents to the ILTA survey was that of procedures for ensuring quality, standards to be upheld or adhered to, as in codes of practice. A second meaning was that of levels of proficiency - what standard have you reached?. A related, third meaning was that contained in the phrase 'standardized test', which typically means a test whose difficulty level is known, which has been adequately piloted and analyzed, the results of which can be compared with those of a working population: standardized tests are typically norm referenced tests. In the latter context 'standards' is equivalent to 'norms'.

In recent years, language testing has sought to establish standards in the first sense (codes of practice) and to investigate whether tests are developed following appropriate professional procedures. Groot argues that the standardization of procedures for test construction and validation is crucial to the comparability and exchangeability of test results across different education settings. Alderson and Buck and Alderson et al. describe widely accepted procedures for test development and report on a survey of the practice of British EFL examining boards. The results showed that current (in the early 1990s) practice was wanting. Practice and procedures among boards varied greatly, yet (unpublished) information was available which could have attested to the quality of examinations.

Exam boards appeared not to feel obliged to follow or indeed to understand accepted procedures, nor did they appear to be accountable to the public for the quality of the tests they produced. Fulcher and Bamford (1996) argue that testing bodies in the USA conduct and report reliability and validity studies partly because of a legal requirement to ensure that all tests meet technical standards. They conclude that British examination boards should be subject to similar pressures of litigation on the grounds that their tests are unreliable, invalid or biased. In the German context, Kieweg (1999) makes a plea for common standards in examining EFL, claiming that

within schools there is little or no discussion of appropriate methods of testing or of procedures for ensuring the quality of language tests (Alderson and Banerjee: 2001).

The purpose of test

Test is used to measure students' mastering with the subject given. Some experts mention the other purpose of test. According to Nurkanca and Sumartana(1986), a test has many purposes. First, is to know how far the result of a programmer applied whether it has reached its goal or not. Second, is to see whether the materials should be re-taught or not. Third, is to get some information about the students' weakness and difficulties in learning about the given materials. Fourth, is to determine the students' achievement and to allow them going through to the grade. Fifth, is to select and group students based on their achievement.

David (1959) conducted six objectives of language testing:

1. To determine readiness for instructional programs.
2. To classify or place individuals in appropriate language classes.
3. To diagnose the individual's specific strengths and weaknesses.
4. To measure aptitude for learning.
5. To measure the extent of student achievement of the instructional goals.
6. To evaluate the effectiveness of instruction.

Characteristic of a Good Test

A test is an important instrument in teaching learning process to measure students' mastery on the materials. To know the affectivities of a test, it has criteria for testing a test. According to Arikunto, there are some criteria of good test; validity, reliability, objectivity, practicality, economy (Brown: 2004).

Validity

A test was classified to be valid if it measures accuracy what it is intended to measure. According to Heaton, validity of a test is the extent to which it measure what it is supposed to measure and nothing else. There are four types of validity ; face validity, content validity, construct validity, and empirical validity (Heaton: 1988).

Reliability

One of the necessary characteristic of good test is reliability. The test was said to be reliable if it is consistent in the measurements. It means that the students must have same mark if the test marked by two or more examiners. Moreover, the reliability of the test was considered a number of factors that may contribute to the unreliability of the test. According to Heaton, the factors affecting the reliability are:

- 1) The extent of the material selected for testing. Reliability is concerned with the size of the test; it is not too long and not too short.
- 2) The administration of the test(Heaton: 1988).
The students or test-takers must have same condition and time limit.
- 3) The instruction. The clarity of the instruction will affect the students' comprehension to answer the test.
- 4) Personal factors, such as motivation and illness.
- 5) Scoring the test. It means that the objective test is more reliable than the subjective test.

There are some methods to estimate reliability. Such as test – retest method, split half, equivalent method, and internal consistency method. Here, the reseacher uses split half method to get reliability because the test did only one times.

This formula is

$$r_{\frac{11}{12}} = \frac{N \sum Y_1 X_1 - (\sum X_1) (\sum Y_1)}{\sqrt{\{(N \sum X_1^2) - (\sum X_1)^2 (N \sum Y_1^2) - (\sum Y_1)^2\}}}$$

After that the result above to corelation with sperman Brown pattern, this formula is :

$$r_{11} = \frac{2 \times \left(r_{\frac{11}{12}} \right)}{1 + \left(r_{\frac{11}{12}} \right)}$$

This is criteria of reliable

0.00-0.20	Not reliable
0.20-0.40	Less Reliable
0.40-0.60	Reliable enough

0.60-0.80	Reliable
0.80-1.00	Very Reliable

Objectivity

According to Arikunto (1993) the test is called objective if it is free from subjective factors which influence the test. Objectivity of a test can be increased by using more objective types test items and the answers are scored according to model answers provided. Arikunto adds that there are two factors that influence the objectivity of a test they are the form of a test and the test scorer.

Practicality

A test is called as practical test if it is easy to do and does not require many equipments and give freedom to the students to do the easier part, easy to score, is completed with clear instructions. Arikunto (1993) stated that practicality of a test deals with a level of difficulties in administering the test itself.

Item Analysis

The purpose of items analysis was to identify the test items whether it is good or not. To know the answer, all items should be identified from the index of difficulty and index discrimination.

METHODOLOGY

Based on the problem statements of this study, the goal of this research is to explain the answer about the language testing technique that used by Sekolah Indonesia Kuala Lumpur and explain how the test valid with the student, so this research use qualitative descriptive study.

This study the writer will use descriptive methodology. This descriptive study is designed to obtain information concerning particular issues and then describe them. Arikunto (1993) states that descriptive research is not meant to test a certain hypothesis, but it only describes the phenomena, situation and condition that occur during the study.

Best moreover divide descriptive research into four parts: document or content analysis study, case study, ethnographic study,

and explanatory observation study. Document or content analysis study is the study which is concern with the explanation of the status of phenomenon at particular time. Case study is the way of organizing social data for the purpose of viewing social reality. Ethnographic study is the process of collecting data on many variables on an extended period of time, in naturalistic setting. Explanatory observation study is the study which seeks to find answers to question through the analysis of variable relationship (Best :1981).

From the statements above, it can be concluded that the study is categorized as document or content analysis study since this study concern about the answer of what the language testing technique that use in Sekolah Indonesia Kuala Lumpur and also this research is concurrent embedded method because it is combine between qualitative approach and quantitative approach to get and analyze that data.

Data Collection Technique and Instrument

In this research, the researcher use study document technique to answer the questions. The teacher made English test, the answer key and Standard of Graduates Competence academic year 2012-2013 are used to answer the validity of the test. The students' answers sheet and the students' scores of the teacher-made English test are used to answer realibility, the index of difficulty and index of discrimination, distracters of the English test. Those instruments are to prove the answer for all questions.

Data Analysis Procedure

In this study, researcher use interview technique because from the researcher point this technique is appropriate to collect the data and this technique is easiest one to know the answer of researcher question. The researcher conduct step in analyzing the data, as follows:

1. Explain the document (English final exam test) in second grades of Sekolah Indonesia Kuala Lumpur.
2. Analyzing the test based on language test technique and measures the test whether valid or not.

Analyzing The Face Validity

Face validity will be high if the students or test takers encounter some or the entire characteristic of good face validity, as follow:

Step	Aspect of test and questions	Explanation
1	Test appearance How is the cover of test? How is the letter used in the test? How is the test layout How is the size of test paper used?	
2	The direction How is the general instruction of the test? How is the specific instrument of the test? How is the instruction for going on to text section in the next page?	
3	Test items types How many types of the test have been chosen? How are the test presented?	

Analyzing the Content Validity

In analyzing the content validity, the researcher want to know accuracy of English test with the indicators of curriculum or Standard Competencies. The researcher collect the data through the following steps:

1. Making a list of the standard competencies, basic competencies, indicators, and learning experience for the tenth grade students of senior high school and the indicators of basic competencies given.
2. Placing each of the test items in the appropriate place with the standard competencies and basic competencies to identify whether or not the standard competencies and basic competencies covered by the final test.

3. Counting the percentage of the test items of every language aspects.
4. Concluding the result of analysis.

ITEM ANALYSIS

Index of Difficulty

The Index of difficulty of an item simply shows how easy or difficult the particular item proved in the test (Heaton : 1988). To analyze the index of difficulty of test items, the researcher takes the following steps:

1. Arranging the students' score from the highest score to the lowest one.
2. Finding the top and the bottom of the students' score, as upper and lower groups. Dividing the scripts in rank order of total score into two groups of equal size, the top half as the upper level and the bottom half as the lower group.
3. Computing the item difficulty by using the formula of by Heaton (1988) bel

$$FV = \frac{R}{N}$$

Where:

FV= Index of difficulty

R = the number of students who answer correctly

N = the number of students who taking the test

Classify the result based on the criteria of Arikunto(1993) , as follow:

- | | |
|--------------------------------|-----------------|
| 1. Test items with 0,00 – 0,30 | Difficult value |
| 2. Test items with 0,31 - 0,70 | Moderate value |
| 3. Test items with 0,71- 1,00 | Easy value |

Analyzing the Index of Discrimination

The Index of Discrimination indicates the extent to which the item discriminates between the testees, separating the more able testees from the less able (Heaton: 1988). To analyze the Index of

Discrimination here use the same steps using in analyze Index of difficulty. Those steps are:

1. Arranging the students' score from the highest score to the lowest one.
2. Finding the top and the bottom of the students' score, as upper and lower groups. Dividing the scripts in rank order of total score into two groups of equal size, the top half as the upper level and the bottom half as the lower group.
3. Calculate the index of discrimination, the researcher used the formula below:

$$D = \frac{\text{Correct U} - \text{Correct L}}{N}$$

Where

- D : The Index of Discrimination
- Correct U : The number of students in upper group who answered the items correctly
- Correct L : The number of students in lower group who answer the items correctly
- N : The number of students taking the test in one group.

Classify the result based on the criteria of Arikunto (1993), as follow:

1. Test items with 0,00 – 0,20 is Poor
2. Test items with 0,21 - 0,40 is Satisfactory
3. Test items with 0,41- 0,70 is Good
4. Test items with 0,71 – 1,00 is Excellent

Analyzing the Effectiveness of Distracter

Besides calculating index of difficulty and discrimination, it also important to analyze the items in very detail, moreover on those which cannot perform as expected. Analyzing the distracter aimed not only to know which items that cannot work properly but also to check why particular test taker failed to answer certain items correctly.

Distracters shave functioned well if these chosen mostly by students from lower level. According to Arikunto (1993), the

distracter which is chosen at least by 5% students from is called good distracter.

In addition, to conduct the effectiveness of distracter the researcher should determine the amount of students from upper and lower level who chosen each options in each item. The researcher also determines the amount of students who do not chose the options at all (omit). However, to ease the analyzing, the researcher used the table below which shows the example of analyzing the Effectiveness of Distracters (Arikunto: 1993)

Item Number	Options	Upper Group	Lower Group	Comment
1	A	1	8	Good
	B*	22	11	Good
	C	1	2	Good
	D	1	4	Good
	O	0	0	NF

FINDINGS

After classifying the students to the upper and the lower group, the next step is analysing the validity and item analysis. The researcher takes two kinds of validity; include face validity and content validity. Item analysis includes index of difficulty, index of discrimination and distracters.

Face validity

To show the result of face validity of English Test for second grade in Sekolah Indonesia Kuala Lumpur, the researcher took two steps. First step was classifying the matter of the test. Second step was analysing test based on criteria in the table below.

First, the test is printed in A4 paper. The test consisted of seven pages. The first page was used to cover. The cover of the test had; a logo of the school, the subject of test, date and time to do the test as well. Second page until six pages was used to reading section that consisted of forty items. The last pages contained essay test, but in this research, the researcher only analyzed the multiple choice test. Second, analyzed the test based on the table below.

The table below shows the result of the analysis of the face validity of the test (Utami: 2013).

Step	Aspect of test and questions	Explanation
1	Test appearance How is the cover of test? How is the letter used in the test? How is the test layout How is the size of test paper used?	The test cover used black colours and suitable font, can be read easily The size of the letter in the test used 12. Can be read easily. The test had good layout. The picture used in the test are understandable The paper in this test used A4 paper
2	The direction How is the general instruction of the test? How is the specific instrument of the test? How is the instruction for going on to text section in the next page?	The general instructions of this test are understandable. This test had no a specific instrument. This test had no instruction going to the next section/ending.
3	Test items types How many types of the test have been chosen? How are the text presented?	This test had 2 types of test. The test had multiple choice test and essay test. This test is quite well presented in the layout or arrangement.

Content Validity

To show the result of the analysis of content validity of the English test for the second grade of high school of Sekolah Indonesia Kuala Lumpur Malaysia, the researcher uses Standard of Graduates competencies of 2012 to know the connection the test with the

standard competencies. The analysing of content validity used table specification

There are seventh columns in that table. The first column contains of standard competence, second column contains of basic competencies, the third column contains of indicators, the forth column contains of learning experience, the fifth column contains of item test that is appropriate with the basic competencies, the next column contains of the number of items test (Σ) and the last column contains of the percentage of total numbers of particular items represent the elated basic competence.

According to J.B Heaton, the test can be said had a good content validity if it covers all the contents as stated in the curriculum. Based on the result of analysing content validity, this test just covers two criteria, the percentage of every aspect of learning content is concluded as follows:

1. There are 45% Or 18 items for reading which focused on narrative, hortatory exposition, and spoof.
2. There are 40% or 16 items for linguistics which focused on simple past, past tense, and adverbs.
3. There are 15 % or 6 items unsuitable because it focused on descriptive and present future tense.

Based on the result above, we can conclude that English test in second grade in Sekolah Indonesia Kuala Lumpur high school Malaysia is good since 85% items test represents all materials. It is more than 50%, according to Bloom if the agreement of the test is 50% or more, it can be concluded that the test had high content validity (Bloom: 1981).

Moreover, there are 6 items or 15% of the test did not cover the materials, they are the items test number 7, 8, 9,10,30,35. Those item are unsuitable with the indicator of standard and basic competencies and were not taught in this semester.

Analysing index of difficulty

To get the data of index of difficulty, the researcher divided the class into 2 groups The first group was upper group, who were

students who get a good score. The second is lower group, who were student who get the bad score.

After the researcher got the data, she did the analysis using formula as follows:

$$FV = \frac{R}{N}$$

Note : FV : index of difficulty
 R : the number of correct answer
 N : the number of students taking test

There are eight columns in the table. First column contained the number of English test items. The second, it was contained the score of the upper group which answer correctly of each English test items. The third contained the score of the lower group who answer correctly of each English test items. The fourth column contained total of upper group and lower group who answer correctly of each items. The fifth column contained the value of index of difficulty. The sixth column contained upper group minus lower group who answer correctly of each items. The seventh column contained the value of index of discrimination. The eight columns contained comment for each item of index difficulty and index discrimination.

The researcher did analysis of the English test in second grade of high School Sekolah Indonesia Kuala Lumpur, the class that the researcher use to collect the data was second grade of Science class, and second grade of Social class. The total numbers of student in those classes were thirty six students. The numbers of student of two classes were taken as a sample of this research. The researcher used those classes because of in that school the second grade presently contains of two classes, Science class and Social class. Each major of the school presently had one class and in each class contained less than 20 students, so the researcher uses those two classes to taken a sample of this research. The students divided into two groups as the upper group consist of eighteen students and lower group consist eighteen students.

After analysing the index of difficulty, the next step is machining the result with the criteria of index of difficulty according to Arikunto. The analysis is organized in the following table.

Criteria of index difficulty

Index of difficulty	Criteria	Item number	Total of item
0,00 – 0,30	Difficult	34,35,40	3
0.31 – 0,70	Moderate	3,4,5,7,8,9,10, 11, 12,13,14,15, 16,17,18,19, 20,21,22,24,25,26,27, 28, 29,31,32, 36,37,38,39.	31
0,71 – 1,00	Easy	1,2,6,23,30,33	6

The table above shows that there are 31 items are moderate level. There are 3 items are difficult level. There are 6 items are easy level. Almost test items are moderate. It means that those items are good to be given to the students. The English test items for second graders high school of Sekolah Indonesia Kuala Lumpur have acceptable index of difficulty.

Index discrimination

Index discrimination is tools to differentiate between students who are in the upper group (achieved well) and the lower group (who did not achieve well). To analyze index discrimination, the researcher arranged student in the upper group and the lower group, same as analysing the index difficulty. After arranging the upper and the lower group then the researcher computed data of the index of discrimination.

There are six eight in the table of analysis index of difficulty and index of discrimination. The first columns contained of the number of items. The second column contained the score of student in upper group who answer correctly of each item. Third column contains the score of the lower group who answer correctly of each item. Fourth column contained of total of the upper group and the lower group who answer correctly each items. Fifth column contained the value of index of difficulty. The sixth column contained the

numbers of students in the upper group minus the number of students in lower group who answer correctly of each items. Seventh column contained the value of index of discrimination. The eight columns contain comment for each item of index difficulty and index discrimination. To calculate the index discrimination for each item number, the formula used:

$$D = \frac{\text{Correct } U - \text{Correct } L}{n}$$

- D : index of discrimination
 Correct U : the number of students in upper group who answer the item correctly
 Correct L : the number of students in lower group who answer the item correctly
 N : number of candidate of one group

Criteria of index discrimination

Index of discrimination	Criteria	Item number	Total of item
0,00 – 0,20	Poor	1,2,4,5,6,7,8,11,12,14,16,18,22,23,24,26,27,30,31,32,33,34,35,39,40	25
0.20 – 0,40	satisfactory	3,9,10,13,15,17,19,20,21,25,28,29,36,37	13
0,40 – 7,00	Good	38	1
0,70 – 1, 00	Excellent	-	-
-0	Wrong	-	-

Based on the table, the result of index of discrimination shows that there are 25 items had poor index of discrimination, there are 13 items had satisfactory, and there are 1 items had good index of discrimination. Almost students index of discrimination are poor. It means that those items are categorized poor. It means that the English test must be revised.

Analyzing the Effectiveness of Distracters

Item distracters are the incorrect options in the multiple choices which distract the testee from the correct answer. A good distracter will attract more students from the lower group than the upper students. Thus, if there are more able students chosen the distracters, it means that the item does not function as expected in it must be revised.

This English test contains of forty items test and each items test had five answer options. This appendix contains of four columns each items. The first column is the number of items test. Second column is the total correct answer of items from the upper group. Third column is the total correct answer from the lower group. The last column is the comment of the distracter.

According to Arikunto (1993), if the distracter was chosen at least by 5% of student who take the test, it is called a good test. ($5\% \text{ from testee} = 5\% \times 60 \text{ students} = 3 \text{ students}$). In this case 5% of the total student is 2 students ($5\% \times 36 \text{ students}$).

The result of distracters shows that most of all distracters had good criteria because the distracters have been chosen by more the lower group than the upper group. So the English test had good distracters.

DISCUSSION

The result of face validity above shows that the test had the criteria of good test. From the cover of the test, it had clear font and colour. The test also had fine letter size to be read. In addition, the test had the acceptable paper size, the test used A4 paper. From the instructions, the instructions are simple and clearly understandable. The first instruction contains date of the test, time to do the test, and how the test must be done. The instruction of each section used unclear instructions. The instruction of each section had items number without explanation about the section. The last criteria are about the kind of the test. The test contain 40 items multiple choice, and 10 essay. From the explanation above the researcher concluded that the English test of second grade of high school Sekolah Indonesia Kuala

Lumpur Malaysia had acceptable quality, not actually well other than acceptable to the students.

Based on the result of content validity above, the test had 85 % items that covered the indicators of Standard of competencies of 2012. It is 15% items that did not cover the indicators of Standard of Graduates Competencies.

The 45% of the test is focused on reading skill. The test had narrative text, hortatory exposition and spoof text. However in this test there is one descriptive test that not thought in this semester.

Then, 40 % from the test is focused on linguistics. There are simple past, past tense, and conjunction. But in this test there is one of items test that use present future that not thought in this semester.

According to Bloom, if the test agreement is 75% or more, then it can be said that the test had high content validity. On the other hand, if agreement is less than 50% the rest is considered having low content validity (Bloom: 1981)

From the explanation above the researcher concluded that the English test that used for second grade Sekolah Indonesia Kuala Lumpur high school Malaysia had good content validity which had 85% covered indicator of Standard Competencies.

Index of difficulty

Based on the table index of difficulty, the result reported that there are 3 items that classified in difficult criteria. In the index difficulty, 3 out of 40 had difficult value had 0,00 until 0,30. These items must be revised because it is too difficult to be done by the students. This test had 6 out of 40 of easy criteria. The easy value of index difficulty is from 0,71 until 1,00. These items are too easy to the students. This items tests it must be revised.

It can be concluded that most of items or 31 out of 40 items are moderate. The moderate criteria had value from 0,31 until 0,70. These items tests were acceptable for the students. This items test did not to be revised.

Index of discrimination

From the result of index discrimination, it was explained that the test have 1 items of the 40 items in good criteria. The good discrimination value is 0, 40 until 0, 70. This item is not to be revised

but it must be put in more items to make a good test. Then, this test had 13 out of 40 items that include in satisfactory discrimination value. These items must be revised because it is not a big number as a poor criteria.

Then the researcher concluded that the test is poor criteria. The tests have 25 out of 40 items. In this discrimination value, the poor criteria had 0, 00 – 0, 19 value. These criteria must be revised because the number of poor criteria is major of the total of test items.

Analyzing of effective distracter

Item distracters are the incorrect options in the multiple choice which is can amuse the student who do the test from the actual answer. A good distracter will attract more students from the lower group than the upper students (Utami: 2013).

In this English test of second grade of Sekolah Indonesia Kuala Lumpur Malaysia had forty multiple choice items. Those are number one until number forty, and each items number was contained of five options. So the item distracters of this test were 160.

From the result above, 7 out of 160 were bad distracters because those item distracters were chosen less that 5% of the total students who take the test. The distracters items must be revised. Besides that, there are 153 out of 160 was good item distracters because the items was chosen by 5% or more of the total of the students who take the test.

In Addition, according to Nurgiyantoro the data for analyzing the effectiveness of distracters in the appendix 5 showed that there are 4 out of 160 non function distracters since none from both the upper group and the lower group of students chosen those distracters. Besides, there are 8 out of 160 distracters categorized as adequate, because they had same amount of voters from the upper and the lower group. Moreover, there are 4 out of 160 malfunction distracters since those items attracts more students in the upper group than students in the lower group, which is good distracters must been chosen by more the lower group than the upper group. These items must be revised. However, there are 144 distracters are good since worked properly to the students. That is concluded that the test had good distracters and not to be revised.

CONCLUSION

From data analysis and discussion in chapter IV, the researcher concluded that English test has been constructed by English teacher in senior high school Sekolah Indonesia Kuala Lumpur Malaysia.

The English final test had been conducted by the English teacher. The test was objective test that consisted of forty multiple choice questions and had five multiple choice objections. From the validity, face validity, English test of the school had acceptable quality, not actually well but acceptable to the students. Then, from content validity, English test of the school had good content validity which had 85% covered indicator of Standard Competencies. In addition from items analysis, index difficulty, most of items or 31 out of 40 items are moderate value. These items tests were acceptable for the students. From item analysis index discrimination, the test is poor criteria. The tests have 25 out of 40 items. These criteria must be revised because the number of poor criteria is major of the total of test items. From item analysis distracters, there are 144 out of 160 distracters are good since worked properly to the students. That is concluded that the test had good distracters and not to be revised.

REFERENCES

- Alderson, J. C. & Banerjee, J. (2001). *Language testing and assessment (Part I)*, United Kingdom: Cambridge University.
- Arifin, Z. (2012). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (1993). *Dasar-dasar Evaluasi Pendidikan*, Jakarta, Bumi Aksara.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. USA: Oxford University Press.
- Brown, H. D. (2001). *Teaching by Principles: An Interactive Approach to Language Pedagogy, Second Edition*. San Francisco: Longman Inc.
- Daryanto. (1999). *Evaluasi Pendidikan*, Jakarta: Rineka Cipta.
- Depdiknas. (2004). *Standard Kompetensi Mata Pelajaran Bahasa Inggris SMP dan Madrasah Tsanawiyah*, Depdiknas: Jakarta,

- Fulcher, G. (2007). *Language testing and assesment*, New York: Routledge
- Heaton, J. B. (1988). *Writing English Language Test*, New York: Longman.
- Hughes, A. (2003) *Testing for Language Teacher*, Cambridge: University Press.
- James, Ayodele, & Oluwatayo. (2012) *Validity and Reliability Issues in Educational Research (Vol 2)*,Nigeria: Institute of Education,Ekiti State University, 2012
- Khoiro, A. (2012). *An analyzed teacher made English try out fr national exam for the third graders of MAN Sidoarjo*, Thesis S1, Surabaya: Perpustakaan IAIN
- Kubiszyn, T. & Borich, G. (2003). *Educational Testing and Measurement* Singapore: John Wiley & Sons, INC.
- Mayangsari, I. M. (2009). *An Analysis of UAS English Test of Second Semester 2008/2009 by Teacher-made English Test in SMA 2 Muhammadiyah Sidoarjo*, Surabaya: Perpustakaan IAIN Sunan Ampel Surabaya,
- Norris, J. M. (2012). Purposeful Language Assessment: Selecting the Right Alternative Test. *English Teaching Forum*, 38(1), pp. 41 – 45. Retreved from: <http://files.eric.ed.gov/fulltext/EJ997530.pdf>
- Nurkanca, W. & Sumartana. (1986). *Evaluasi Pendidikan*, Surabaya: Usaha Nasional.
- Sudijono, A. (1996) *Pengantar Evaluasi Pendidikan*. Jakarta: PT. Raja Grafindo Persada.
- Tambini, R. F. (1999). Aligning learning activities and assessment strategies in the ESL classroom. *The Internet TESL Journal*, V(9). Retreved from: <http://iteslj.org/Articles/Tambini-Aligning.html>
- Utami, S. 2014. *An analysis teacher made English UKK test for academic years 2012 – 2013 for seventh graders of Muhammadiyah 9 surabaya*, Thesis S1, (Unpublished)